

# Multi-scale Context Intertwining for Semantic Segmentation

Di Lin<sup>1</sup>, Yuanfeng Ji<sup>1</sup>, Dani Lischinski<sup>2</sup>, Daniel Cohen-Or<sup>1,3</sup>,  
and Hui Huang<sup>1</sup>(✉)

<sup>1</sup> Shenzhen University, Shenzhen, China  
ande.lin1988@gmail.com, jyuanfeng8@gmail.com, cohenor@gmail.com,  
hhzhiyan@gmail.com

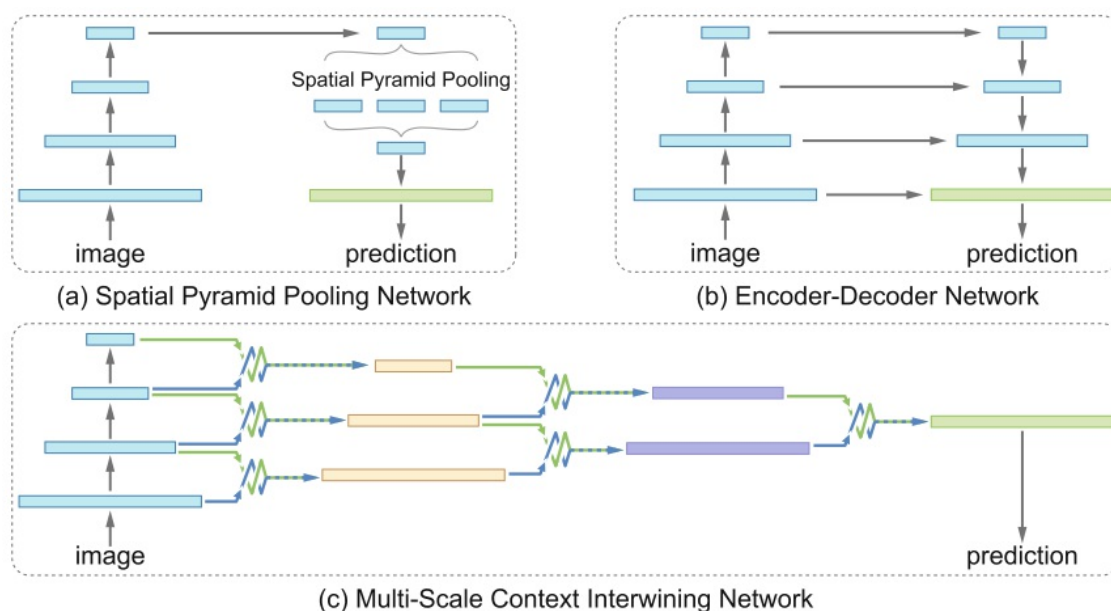
<sup>2</sup> The Hebrew University of Jerusalem, Jerusalem, Israel  
danix3d@gmail.com

<sup>3</sup> Tel Aviv University, Tel Aviv, Israel

**ECCV2018**

## 目的：

本文提出一种新的特征融合的方法，与前向链接不同，而使用 LSTM 横向地对每一层的网络输出做融合，相邻尺度的语义相关性强，所以它们之间互相交织可以得到更好的结果。

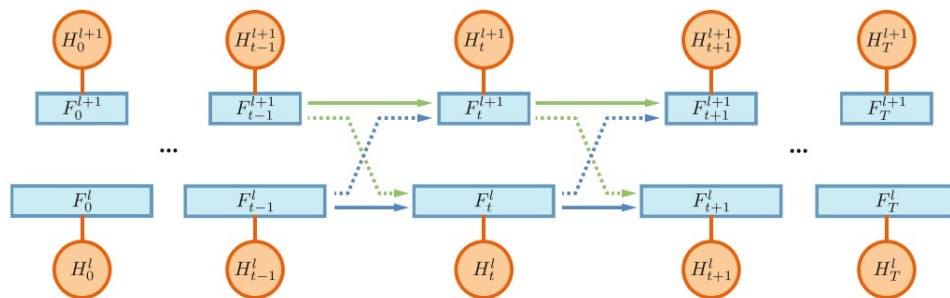


## 方法：

多尺度上下文交织：由两个 LSTM 单元构成双向的信息流，在  $l$  层和  $l+1$  层特征图之间，互相提取有用的信息给对方；然后将  $l+1$  层的低分辨率特征图做

转置卷积使得分辨率与 l 层的相同；二者相加得到输出 Q；接着送到下一阶段的交织,整个过程持续 T 步。

由于 LSTM 的记忆性,使得每一层的特征图信息都能被送到最后生成的特征图,即包含了局部和全局的多尺度特征得到了融合。



双向连接：

首先将特征图分成一个个不重合的区域,用的是传统基于边缘的方法；将感受野中心位于区域内的每个神经元聚合,再将目标区域 S 及其相邻区域的信息聚合,送到另一级特征图中；这个过程是双向的。

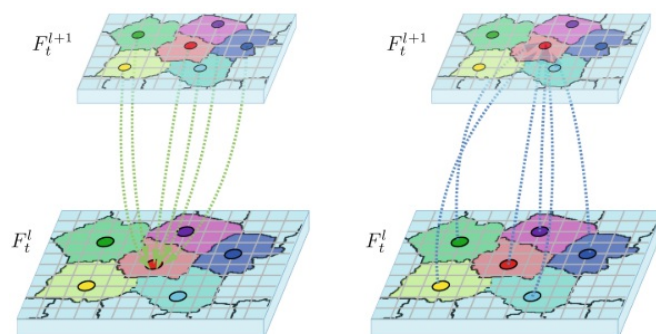
区域特征：

$$R_{n,t}^l = \sum_{(h,w) \in \Phi(S_n)} F_t^l(h,w),$$

聚合的相邻区域信息的全局特征：

$$M_{n,t}^l = \sum_{S_m \in \mathcal{N}(S_n)} R_{m,t}^l.$$

得到的 R、M 作为下一个节点的输入。



LSTM 的内部计算：

根据当前输入和上个节点的输出分别计算输入门(i)、忘记门(f)、输出门(o)、和输入的变换(c),按位置的矩阵乘法计算得到当前的输出 A,即融合之后的特征,再加上当前层级的特征得到交织后的特征图。

$$\begin{aligned}
G_{i,t}^{l \rightarrow l+1}(h, w) &= \sigma(W_i^{l+1} * F_t^{l+1}(h, w) + W_{s,i}^l * R_{n,t}^l + W_{a,i}^l * M_{n,t}^l + b_i^{l+1}), \\
G_{f,t}^{l \rightarrow l+1}(h, w) &= \sigma(W_f^{l+1} * F_t^{l+1}(h, w) + W_{s,f}^l * R_{n,t}^l + W_{a,f}^l * M_{n,t}^l + b_f^{l+1}), \\
G_{o,t}^{l \rightarrow l+1}(h, w) &= \sigma(W_o^{l+1} * F_t^{l+1}(h, w) + W_{s,o}^l * R_{n,t}^l + W_{a,o}^l * M_{n,t}^l + b_o^{l+1}), \\
G_{c,t}^{l \rightarrow l+1}(h, w) &= \tanh(W_c^{l+1} * F_t^{l+1}(h, w) + W_{s,c}^l * R_{n,t}^l + W_{a,c}^l * M_{n,t}^l + b_c^{l+1}), \\
C_{t+1}^{l \rightarrow l+1}(h, w) &= G_{f,t}^{l \rightarrow l+1}(h, w) \odot C_t^{l \rightarrow l+1}(h, w) + G_{i,t}^{l \rightarrow l+1}(h, w) \odot G_{c,t}^{l \rightarrow l+1}(h, w), \\
A_{t+1}^{l \rightarrow l+1}(h, w) &= \tanh(G_{o,t}^{l \rightarrow l+1}(h, w) \odot C_{t+1}^{l \rightarrow l+1}(h, w)), \\
F_{t+1}^{l+1}(h, w) &= F_t^{l+1}(h, w) + A_{t+1}^{l \rightarrow l+1}(h, w),
\end{aligned} \tag{5}$$

## 总结：

---

### Algorithm 1. Multi-Scale Context Intertwining

---

- 1: **Input:**
    - 1) the number of stages  $T$  for each phase of the context intertwining;
    - 2) a set of convolutional feature maps  $F = \{F^l\}$ , where  $l = 1, \dots, L$ ;
    - 3) the trained parameter set  $\{(P^{l \rightarrow l+1}, P^{l+1 \rightarrow l})\}$ .
  - 2: **Initialization:**
    - 1) a total  $K$  phases for the context intertwining, where  $K = L - 1$ ;
    - 2) a set  $Q = \{Q_k\}$ , where  $Q_0 = \{Q_0^l | Q_0^l = F^l\}$ ; and  $Q_k = \emptyset, k = 1, \dots, K$ ;
    - 3) a set of cell states  $\{(C^{l \rightarrow l+1}, C^{l+1 \rightarrow l})\}$ , where  $C^{l \rightarrow l+1}, C^{l+1 \rightarrow l} = 0$ .
  - 3: **for**  $k = 1 \rightarrow K$  **do**
  - 4:   **for**  $l = 1 \rightarrow |Q_{k-1}| - 1$  **do**
  - 5:      $\{Q_k^l, C_T^{l \rightarrow l+1}, C_T^{l+1 \rightarrow l}\} = \mathcal{L}(Q_{k-1}^l, Q_{k-1}^{l+1}, C^{l \rightarrow l+1}, C^{l+1 \rightarrow l}, P^{l \rightarrow l+1}, P^{l+1 \rightarrow l}, T)$
  - 6:      $Q_k = Q_k \cup \{Q_k^l\}, (C^{l \rightarrow l+1}, C^{l+1 \rightarrow l}) = (C_T^{l \rightarrow l+1}, C_T^{l+1 \rightarrow l})$
  - 7:   **end for**
  - 8: **end for**
  - 9: **Output:** the segmentation feature map  $Q_K^1$ .
- 

网络每一层都有一个单独的 LSTM 单元，两层特征的交织过程是用上述 LSTM 迭代多次，将一个特征图某个区域内的局部和全局信息融合到另一个特征图。相邻两层特征图两两交织之后用转置卷积合成一个特征图，直到最后剩下一个。